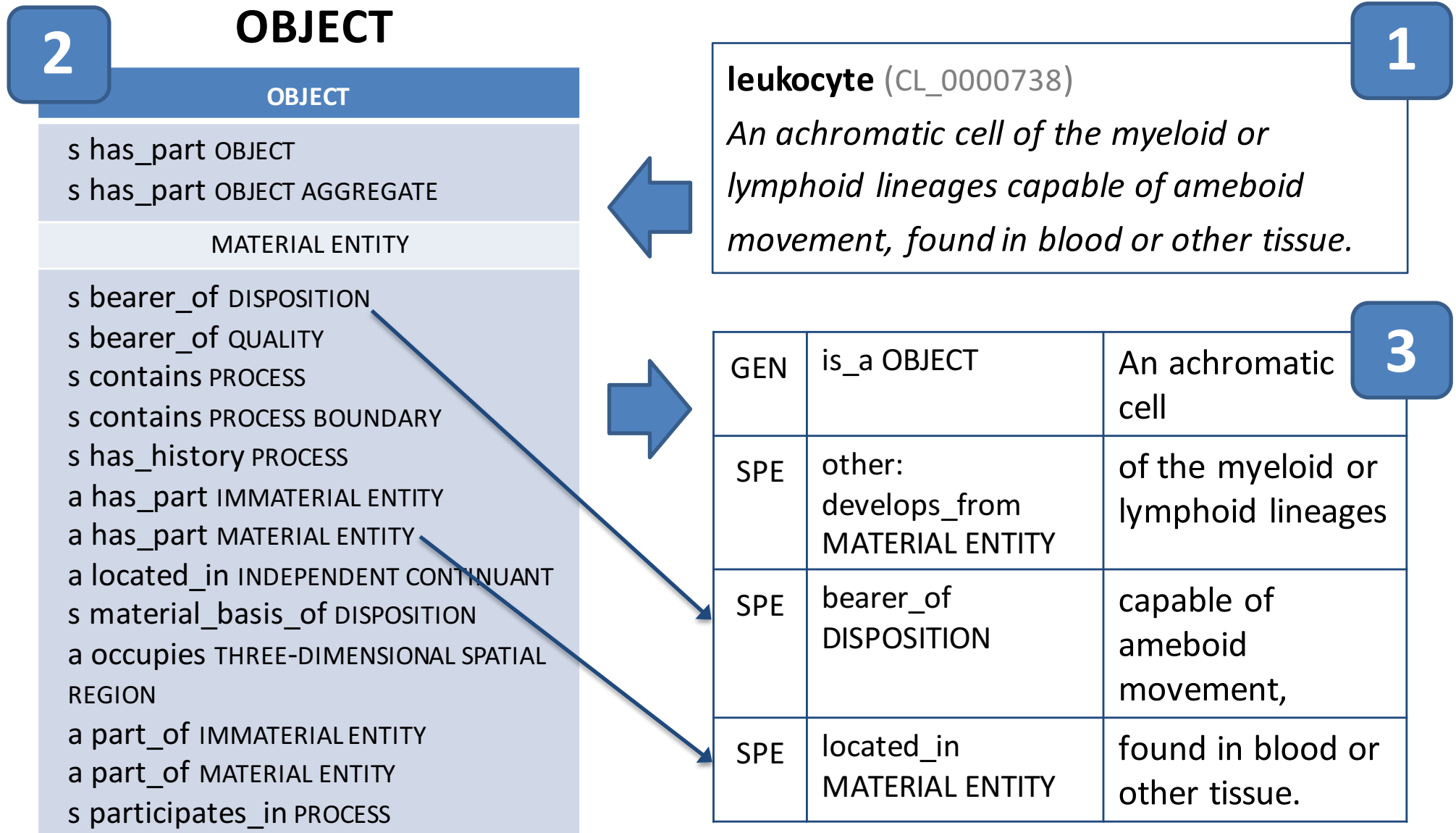# Semi-Automatic Mapping of WordNet to Basic Formal Ontology

Selja Seppälä, Amanda Hicks, and Alan Ruttenberg

Global WordNet Conference
Bucharest, Romania
January 30, 2016

Background (1)

OBJECT

2

OBJECT
s has_part OBJECT
s has_part OBJECT AGGREGATE
MATERIAL ENTITY
s bearer_of DISPOSITION
s bearer_of QUALITY
s contains PROCESS
s contains PROCESS BOUNDARY
s has_history PROCESS
a has_part IMMATERIAL ENTITY
a has_part MATERIAL ENTITY
a located_in INDEPENDENT CONTINUANT
s material_basis_of DISPOSITION
a occupies THREE-DIMENSIONAL SPATIAL REGION
a part_of IMMATERIAL ENTITY
a part_of MATERIAL ENTITY
s participates_in PROCESS

1

leukocyte (CL_0000738)

An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement, found in blood or other tissue.

3

GEN	is_a OBJECT	An achromatic cell
SPE	other: develops_from MATERIAL ENTITY	of the myeloid or lymphoid lineages
SPE	bearer_of DISPOSITION	capable of ameboid movement,
SPE	located_in MATERIAL ENTITY	found in blood or other tissue.

Background (2)

- BFO also used in the biomedical field and increasingly in other domains

- Existing mappings of WN to upper-level ontologies
  - WN-DOLCE (Gangemi et al., 2010)
  - WN-KYOTO (Laparra et al., 2012)
  - WN-SUMO (Niles and Pease, 2003; Pease and Fellbaum, 2010)

- No lexico-semantic resource available for the Basic Formal Ontology (BFO)

➜ Create a BFO-compliant lexical resource

BFO 2.0

- A domain-neutral formal upper-level ontology (Smith et al., 2012)

- Represents the types of things that exist in the world and relations between them

- Serves as an integration hub for mid-level and domain-specific ontologies, which thus become interoperable (Smith and Ceusters, 2010)

- Previous versions (BFO 1.0 and BFO 1.1) have been mapped to BFO 2.0 (Seppälä et al., 2014)

Goal & hypotheses

- Semi-automatically mapping WordNet 3.0 to BFO 2.0

- A large portion of WN synsets, especially nouns and verbs, can be semi-automatically mapped to BFO

- Exploiting existing mapping between WN and the KYOTO ontology

The KYOTO ontology

- For representing domain-specific terms in a computer-tractable axiomatized formalism to allow machines to reason over texts in natural language (Vossen *et al.*, 2010)

- Links WordNets of different languages to ontology classes via a mapping of the English WordNet to KYOTO

- Subdivided into three layers

- Includes DOLCE (an upper-level ontology similar to BFO)

KYOTO ontology's three layers
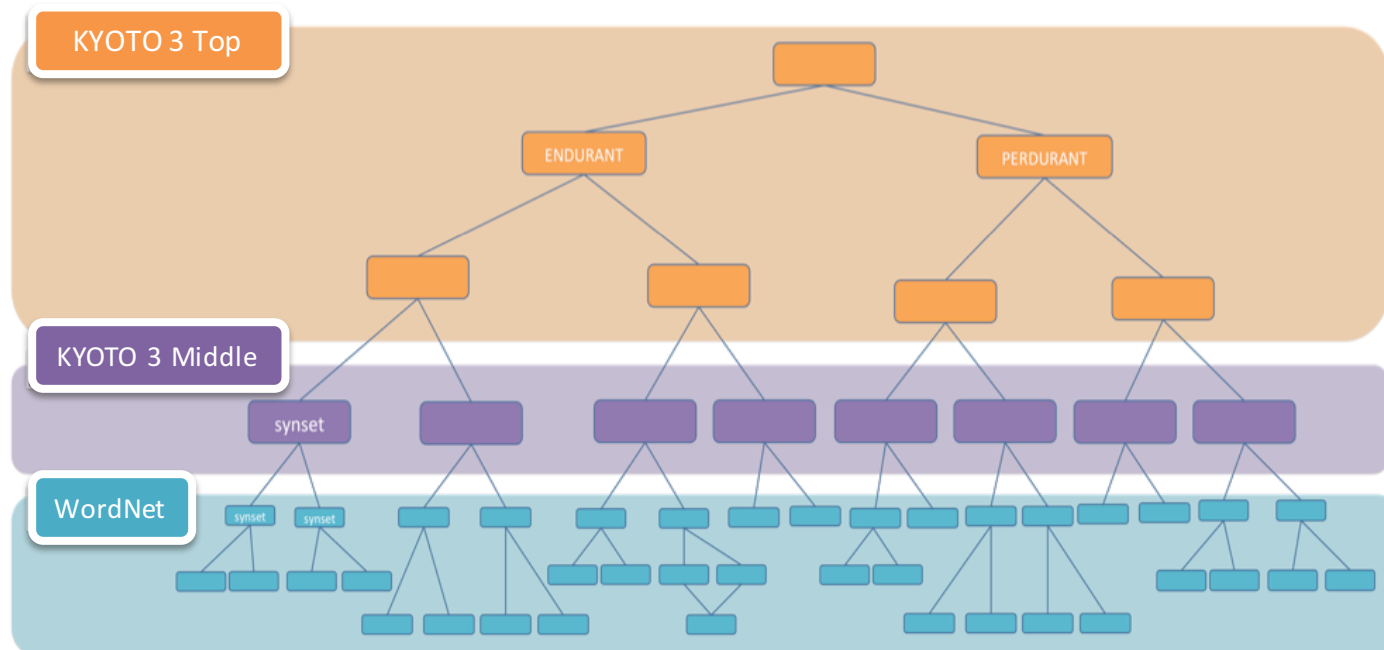
Layer 1:
KYOTO 3 Top Ontology
Includes the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE-Lite-Plus, version 3.9.7)

Layer 2:
KYOTO 3 Middle Ontology
Noun and verb synsets constituting a set of Base Concepts (BCs)

Layer 3
WordNet
WN synsets containing domain-specific classes (e.g. from the environmental domain)

Relevant ontological characteristics

- DOLCE and BFO share relevant characteristics
  - Domain neutrality
  - Strict hierarchical *is_a* taxonomy
  - Bi-partition into CONTINUANTS (DOLCE 'endurants') and OCCURRENTS (DOLCE 'perdurants')
  - Distinction between independent and dependent entities
- Can be loosely mapped exploiting previous mappings

Method overview

KYOTO

BFO
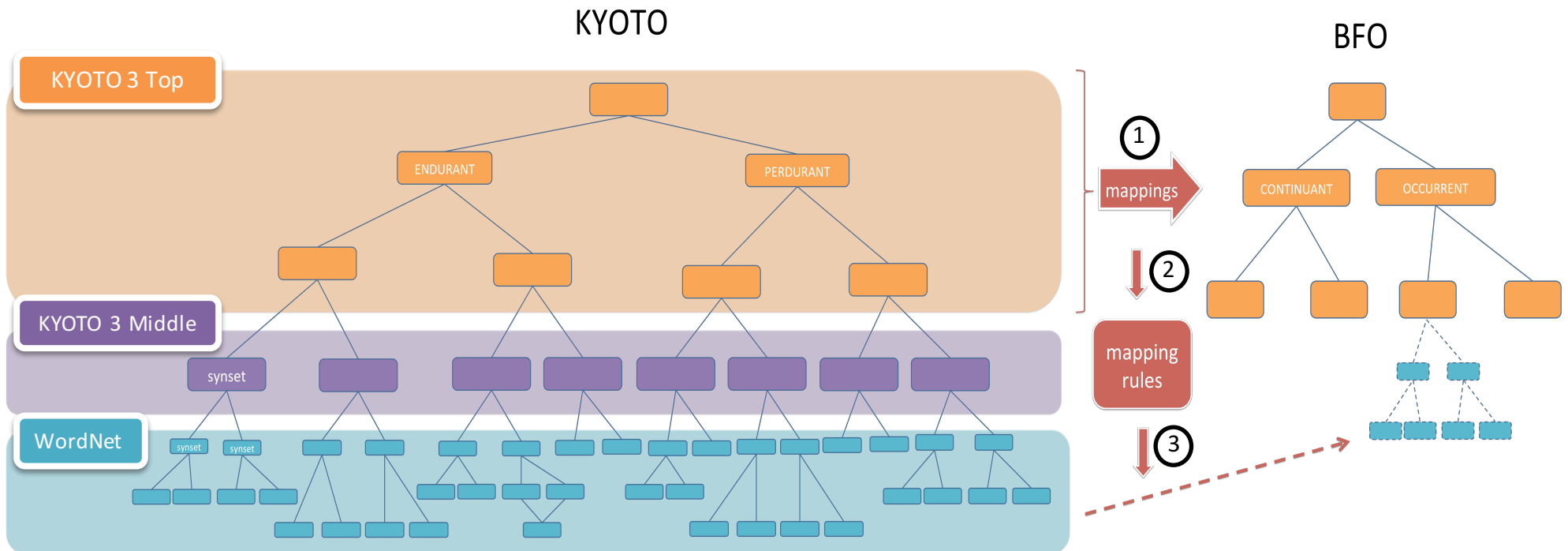
① Mappings from DOLCE to BFO 2.0

② Ruleset mapping KYOTO types to BFO based on ①

③ Get synsets, their base concepts and KYOTO types, and apply rules to map WordNet to BFO

Implementation

immunity.n.02

Input

'Kyoto#condition__status-eng-3.0-13920835-n',
'Kyoto#state-eng-3.0-00024720-n',

'ExtendedDnS.owl#situation',
'ExtendedDnS.owl**#non-agentive-social-object**',
'ExtendedDnS.owl#social-object',
'DOLCE-Lite.owl#non-physical-object',
'DOLCE-Lite.owl#non-physical-endurant',
'DOLCE-Lite.owl#endurant',
'DOLCE-Lite.owl#spatio-temporal-particular',
'DOLCE-Lite.owl#particular'

Program tests if string in rules matches element in list

. . .

'**#non-agentive-social-object** > disposition'
'accomplishment > process'
'noun.act > process'
. . .

Output

immunity.n.02 >
DISPOSITION

Evaluation

- Synset datasets
  - 'medicine sample': 106 nouns & verbs marked 'medicine'
  - 'POS-sample': 100 nouns & 100 verbs extracted randomly

- Rulesets
  - 1st ruleset
    - Created using DOLCE/KYOTO/WN lexname mappings to BFO
    - Tested on 'medicine sample'
  - 2nd ruleset
    - Tuned on 'medicine sample'
    - Tested on 'POS-sample'

Goldstandard

- Manually created by BFO experts

- Intuitive categorization criterion
  Assign the most specific BFO type of which the referent of the synset is a subtype.

- Example
  'the synset `immunity.n.02` refers to a subtype of the BFO type DISPOSITION'

Baseline

- ## WN nouns
  - noun.tops manually mapped to BFO 2.0
  - Propagating mappings downwards to WN synsets

- ## WN verbs
  - Automatically mapped to BFO 2.0 PROCESS

- ## Limitation
  - Not always mappable to lower-level BFO categories

Results: Correct mappings

% of WN-BFO mappings	medicine n-v sample								
	baseline			first ruleset			new ruleset		
	n	**v**	**total**	**n**	**v**	**total**	**n**	**v**	**total**
correct	55	100	76	70	85	77	72	100	85
partial	17	0	9	0	12	6	0	0	0
incorrect	28	0	15	28	2	16	26	0	14
no mapping	0	0	0	2	0	1	2	0	1
total	100	100	100	100	100	100	100	100	100

- ↗ in 'medicine sample', but 2nd ruleset tuned on it
- ↘ in 'POS-sample' due to incorrect and non-mapping verbs
- BUT ↗ using baseline rule
- Slight ↗ for nouns

% of WN-BFO mappings	pos sample					
	baseline			new ruleset		
	n	**v**	**total**	**n**	**v**	**total**
correct	41	99	70	42	86	64
partial	25	0	12.5	0	0	0
incorrect	34	1	17.5	53	7	30
no mapping	0	0	0	5	7	6
total	100	100	100	100	100	100

Prospective performance

WN-BFO mappings	n	v	total	%
correct	42	99	141	70.5
partial	0	0	0	0
incorrect	53	1	54	27
no match	5	0	5	2.5
total	100	100	200	100

Results: Incorrect mappings

% of WN-BFO mappings	medicine n-v sample								
	baseline			first ruleset			new ruleset		
	n	v	total	n	v	total	n	v	total
correct	55	100	76	70	85	77	72	100	85
partial	17	0	9	0	12	6	0	0	0
incorrect	28	0	15	28	2	16	26	0	14
no mapping	0	0	0	2	0	1	2	0	1
total	100	100	100	100	100	100	100	100	100

- Slight ↘ in 'medicine sample' (n: ↘; v: ↗)
- = in 'POS-sample' (n: ↘; v: ↗)
- BUT ↘ using baseline rule
- No more partial mappings

% of WN-BFO mappings	pos sample					
	baseline			new ruleset		
	n	v	total	n	v	total
correct	41	99	70	42	86	64
partial	25	0	12.5	0	0	0
incorrect	34	1	17.5	53	7	30
no mapping	0	0	0	5	7	6
total	100	100	100	100	100	100

Prospective performance

WN-BFO mappings	n	v	total	%
correct	42	99	141	70.5
partial	0	0	0	0
incorrect	53	1	54	27
no match	5	0	5	2.5
total	100	100	200	100

Results: No mappings

% of WN-BFO mappings	medicine n-v sample								
	baseline			first ruleset			new ruleset		
	n	**v**	**total**	**n**	**v**	**total**	**n**	**v**	**total**
correct	55	100	76	70	85	77	72	100	85
partial	17	0	9	0	12	6	0	0	0
incorrect	28	0	15	28	2	16	26	0	14
no mapping	0	0	0	2	0	1	2	0	1
total	100	100	100	100	100	100	100	100	100

- None in baseline BUT 16% of nouns mapped to ENTITY
- ↗ in 'POS-sample' mostly due to verbs
- BUT ↘ using baseline rule

% of WN-BFO mappings	pos sample					
	baseline			new ruleset		
	n	**v**	**total**	**n**	**v**	**total**
correct	41	99	70	42	86	64
partial	25	0	12.5	0	0	0
incorrect	34	1	17.5	53	7	30
no mapping	0	0	0	5	7	6
total	100	100	100	100	100	100

Prospective performance

WN-BFO mappings	n	v	total	%
correct	42	99	141	70.5
partial	0	0	0	0
incorrect	53	1	54	27
no match	5	0	5	2.5
total	100	100	200	100

General observations

- Verbs better covered than nouns
  - Nouns refer to a wider array of BFO categories

- Verbs best covered by baseline rule
  - Most verbs refer to subtypes of BFO PROCESS

- Nouns best covered by rulesets
  - Rules allow mappings to lowest BFO categories

Ontological challenges

- Ontological distinctions not captured in WN
  - Rigid vs. non-rigid properties
  - WN: 'carrier.n.09' ➔ BFO: OBJECT or ROLE?
  - ➔ Rule where WN: 'noun.person' ➔ BFO: ROLE
- Hierarchichal discrepancies (hyponymy vs. 'is_a')
  - WN not ontologically precise
  - WN: 'symptom.n.01' & 'sign.n.06' hyponyms of 'cognition.n.01'
  - ➔ Use semi-automatic method to ontologically evaluate WN's hierarchy (e.g., Rudify)
  - ➔ Refine mapping rules iteratively

Challenges: One-to-many mappings

- From DOLCE types
  - DOLCE: 'feature' ➜ BFO: SITE or FIAT OBJECT PART?
  - ➜ Further disambiguation step required

- From WN synsets
  - Systematic polysemy
  - WN: 'carpet_beetle.n.01' ➜ BFO: OBJECT or OBJECT AGGREGATE?
  - ➜ Further investigation needed

Challenges: Non-mapping cases

- From DOLCE types
  - DOLCE: 'abstract' ➔ no BFO type

  ➔ Test new rules that might work for some cases

- From WN synsets
  - Non-existent entity types
  - WN: 'mythical_creature.n.01' ➔ no BFO type

  ➔ Problem for automation

Future work

- Examining results further

- Testing, extending & refining mapping rules
  - Mapping KYOTO BCs to BFO and propagating downwards
  - Creating new mapping rules using
    - WN-SUMO mappings
    - Ontologies extending BFO

- Processing systematic polysemy

- Processing adjectives in terms of BFO types

Conclusion

- A method to semi-automatically map WordNet 3.0 synsets to BFO 2.0 types via the KYOTO ontology

  ➔ Identifying challenges

  ➔ Getting a sense of performance

- Encouraging preliminary results

- More work needed to see if method scales to the full WordNet

  ➔ Reduces manual work

- Challenge: Providing BFO-compliant interpretations of unmatched WN synsets

seljamar@buffalo.edu

aehicks@ufl.edu

alanruttenberg@gmail.com

THANK YOU

References (1)

Robert Arp, Barry Smith, and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Aldo Gangemi, Roberto Navigli, and Paola Velardi. 2003. *The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 820–838.

Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. 2010. "Interfacing WordNet with DOLCE: towards OntoWordNet". In Chu-ren Huang, Nicoletta Calzolari, and Aldo Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*, pages 36–52. Cambridge University Press.

Pierre Grenon. 2003. *BFO in a Nutshell: A Bi-categorial Axiomatization of BFO and Comparison with DOLCE*. IFOMIS Report 06/2003. Technical report, Institute for Formal Ontology and Medical Information Science (IFOMIS), University of Leipzig, Leipzig, Germany.

Nicola Guarino and Christopher Welty. 2002. "Evaluating Ontological Decisions with OntoClean". *Commun. ACM*, 45(2):61–65, February.

Axel Herold, Amanda Hicks, German Rigau, and Egoze Laparra. 2009. *Central Ontology Version - 1 Deliverable 6.2*. Technical report.

Amanda Hicks and Axel Herold. 2009. "Evaluating ontologies with Rudify". In Jan L. G. Dietz, editor, *Proceedings of the 2nd International Conference on Knowledge Engineering and Ontology Development (KEOD'09)*, pages 5–12. INSTICC Press.

References (2)

Zubeida Casmod Khan and C. Maria Keet. 2013. "Addressing issues in foundational ontology mediation". In *Proceedings of KEOD'13*, pages 5–16, Vilam- oura, Portugal, September 19–22. SCITEPRESS.

Egoitz Laparra, German Rigau, and Piek Vossen. 2012. "Mapping WordNet to the Kyoto ontology". In *LREC*, pages 2584–2589.

George A Miller. 1995. "WordNet: a lexical database for English". *Communications of the ACM*, 38(11):39–41.

I. Niles and A. Pease. 2003. "Linking Lexicons and Ontologies: Mapping Wordnet to the Suggested Upper Merged Ontology". In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416. Citeseer.

Adam Pease and Christiane Fellbaum. 2010. "Formal ontology as interlingua: The SUMO and WordNet linking project and global WordNet". In Churen Huang, Nicoletta Calzolari, and Aldo Gangemi, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge Univer- sity Press.

Selja Seppälä, Barry Smith, and Werner Ceusters. 2014. "Applying the realism-based ontology-versioning method for tracking changes in the basic formal ontology". In *8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*, Rio de Janeiro, Brazil.

Selja Seppälä. 2015a. "Mapping WordNet to the Basic Formal Ontology using the KYOTO ontology". In *Proceedings of ICBO 2015*.

Selja Seppälä. 2015b. "An ontological framework for modeling the contents of definitions". *Terminology*, 21(1):23–50.

A. Patrice Seyed. 2009. "BFO/DOLCE primitive relation comparison". In *Nature Precedings*.

Barry Smith and Werner Ceusters. 2010. "Ontological Realism: A Methodology for Coordinated Evolution of Scientific Ontologies". *Applied Ontology*, 5:139–188.

References (3)

Barry Smith, Mauricio Almeida, Jonathan Bona, Mathias Brochhausen, Werner Ceusters, Melanie Courtot, Randall Dipert, Albert Goldfain, Pierre Grenon, Janna Hastings, William Hogan, Leonard Jacuzzo, Ingvar Johansson, Chris Mungall, Darren Natale, Fabian Neuhaus, Anthony Petosa Robert Rovetto, Alan Ruttenberg, Mark Ressler, and Stefan Schulz, 2012. *Basic Formal Ontology 2.0: Draft Specification and User's Guide*, July.

Andrew D. Spear, 2006. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany.

Lynda Temal, Arnaud Rosier, Olivier Dameron, and Anita Burgun. 2010. "Mapping BFO and DOLCE". *Studies In Health Technology And Informatics*, 160(Pt 2):1065–1069.

Piek Vossen, German Rigau, Eneko Agirre, Aitor Soroa, Monica Monachini, and Roberto Bartolini. 2010. "KYOTO: an open platform for mining facts". In *Proceedings of the 6th Workshop on Ontologies and Lexical Resources*, pages 1–10.